

Implementasi Algoritma C5.0 Dan K-Medoids

by Galih Wasis Wicaksono

Submission date: 02-Apr-2020 10:13AM (UTC+0700)

Submission ID: 1287506626

File name: 696-1653-1-PB.pdf (613.75K)

Word count: 4057

Character count: 26280

Implementasi Algoritma C5.0 Dan K-Medoids Untuk Klasterisasi Ibu Hamil Beresiko Tinggi

Muhammad Riadi¹, Yufis Azhar², Galih Wasis Wicaksono³

^{1,2,3}Teknik Informatika/Universitas Muhammadiyah Malang
mriadi38@gmail.com¹, yifis.az@gmail.com², galih.w.w@umm.ac.id³

Abstrak

Mengingat permasalahan angka kematian ibu dan balita, dan tingginya angka kehamilan beresiko tidak hanya masalah dunia kesehatan saja. Peranan Teknologi informasi yang berkembang sangat pesat dapat dijadikan solusi terhadap permasalahan kehamilan beresiko. Terlebih lagi, computer sering kali berubah fungsi untuk menggantikan pekerjaan manusia yang bersifat rutinitas maupun pengambilan keputusan. Maka untuk mengatasi kasus ini akan diterapkan dua algoritma yaitu: (a) Algoritma Decision Tree C5.0, (b) K-Medoids Clustering. Metode Commercial Version 5.0 (C5.0) untuk mengolah variabel-variabel analisa yang digunakan. Penggunaan C5.0 pada kasus ini untuk melakukan seleksi atribut sehingga menghasilkan fitur yang sangat berpengaruh. Setelah melakukan seleksi fitur data yang baru akan dikelompokkan menggunakan K-Medoids untuk di analisa agar dapat dijadikan acuan untuk penanganan pada kasus ini. Penerapan kedua metode ini juga agar keputusan yang nanti diambil lebih tepat sasaran untuk mengurangi atau mengatasi masalah kehamilan ibu yang beresiko tinggi.

Kata Kunci: Kematian Ibu dan Bayi, Kehamilan Beresiko, C5.0, K-Medoids

Abstract

Considering the problem of maternal and under-five mortality rates, and the high number of pregnancies at risk is not just a matter of the health world. The role of information technology that is developing very rapidly can be used as a solution to the problem of risky pregnancy. What's more, computers often change functions to turn off routine human work and decision making. Then to overcome this case two algorithms will be applied namely: (a) Decision Tree C5.0 Algorithm, (b) K-Medoids Clustering. Commercial Version 5.0 (C5.0) method for processing the analysis variables used. The use of C5.0 in this case is for attribute selection so that it produces very powerful features. After doing the selection of new data features will be grouped using K-Medoids for analysis so that they can be used as a reference for handling this case. The application of these two methods is also so that the decisions that are made later are more targeted to reduce or overcome the problem of high-risk maternal pregnancy.

Keywords: Maternal and Infant Mortality, Pregnancy At Risk, C5.0, K-Medoids

1. Pendahuluan

Pembangunan kesehatan Dalam mengujudkan tujuan nasional bangsa Indonesia sesuai Pembukaan UUD 1945, maka perlunya kesadaran, kemauan, kemampuan hidup sehat bagi setiap orang agar terwujudnya kemakmuran masyarakat. Kesehatan ibu dan bayi sangat berpengaruh terhadap kesehatan masyarakat. Kesehatan masyarakat dapat meningkat jika menurunnya angka kematian ibu dan bayi, resiko kematian bayi dan balita dapat diatasi jika meningkatkan pertolongan persalinan oleh tenaga kesehatan dimana persalinan oleh tenaga kesehatan persalinan tersebut dilakukan oleh tenaga medis [4]. Menurut data SDKI pada tahun 2016, Angka Kematian Ibu (AKI) saat ini diperkirakan mencapai 300 per 100.000 lebih kelahiran hidup. Fakta ini menunjukkan adanya lonjakan yang sangat signifikan dari hasil SDKI tahun 2012 yaitu sebesar 200 per 100.000 lebih kelahiran hidup [5].

Ini menegaskan bahwa Indonesia, dari angka kematian ibu yang dilaporkan ada 400 ribu ibu meninggal perbulannya dan hampir dari 15 ibu meninggal setiap harinya. Artinya ada masalah dalam program kesehatan ibu dan anak saat ini [6]. Beliau juga mengungkapkan, sebab tertingginya kematian ibu di tahun 2016, sebesar 32% diakibatkan perdarahan. Sementara 26% nya diakibatkan hipertensi yang mengakibatkan terjadinya kejang, dan ibu hamil sangat

rentan terhadap keracunan sehingga menyebabkan ibu meninggal. Dan yang menyebabkan kematian pada bayi baru lahir [6][7].

Mengingat permasalahan ini tidak hanya masalah dalam dunia kesehatan maka peranan teknologi informasi yang berkembang sangat pesat dapat dijadikan solusi terhadap permasalahan kehamilan beresiko [10][11]. Terlebih lagi, computer sering kali berubah fungsi untuk menggantikan pekerjaan manusia yang bersifat rutinitas maupun pengambilan keputusan. Maka untuk mengatasi kasus ibu hamil beresiko akan diterapkan dua algoritma yaitu: (a) Algoritma *Decision Tree* C5.0, (b) *K-Medoids Clustering*. Metode *Commercial Version 5.0* (C5.0) untuk mengolah variabel-variabel analisa yang digunakan.

Penggunaan C5.0 pada kasus ini untuk melakukan seleksi atribut sehingga menghasilkan fitur yang sangat berpengaruh. Setelah melakukan seleksi fitur data yang baru akan dikelompokkan menggunakan *K-Medoids* untuk di analisa agar dapat dijadikan acuan untuk penanganan pada kasus ini. Penerapan kedua metode ini juga agar keputusan yang nanti diambil lebih tepat sasaran untuk mengurangi atau mengatasi masalah kehamilan ibu yang beresiko tinggi.

2. Kajian Pustaka

2.1 Kehamilan Beresiko

Kehamilan Resiko Tinggi adalah kasus kehamilan yang di dalamnya kehidupan atau kesehatan si ibu atau janin dalam keadaan bahaya akibat gangguan dari berbagai aspek kehamilan yang kebetulan atau unik. Ukuran risiko kehamilan ini dapat dituangkan dalam bentuk angka atau yang biasa disebut dengan SKOR. Digunakan bilangan bulat di bawah 10, sebagai angka dasar 2, 4 dan 8 dari tiap faktor sebagai pembeda risiko yang rendah, risiko tinggi, risiko sangat tinggi. Dapat dengan jelasnya dilihat berdasarkan jumlah skor kehamilan dibagi tiga kelompok:

1. Kehamilan Risiko Rendah (KRR) dengan jumlah skor terendah 2

Kehamilan 8 yang tidak memiliki masalah atau faktor resiko, fisiologis sampai dengan fase persalinan yang normal dengan ibu dan bayi hidup sehat.

2. Kehamilan Risiko Tinggi (KRT) dengan jumlah skor sedang 6-10

Pada kehamilan ini memiliki satu resiko atau lebih, baik dari pihak ibu maupun janinnya yang memberi dampak kurang baik bagi ibu maupun janinnya, memiliki risiko kegawatan tetapi tidak darurat dan lebih bisa untuk diatasi tenaga medis.

3. Kehamilan Risiko Sangat Tinggi (KRST) dengan jumlah skor ≥ 12

Pada kehamilan ini memiliki lebih banyak resiko dan biasanya dibarengi dengan berbagai faktor pendukung, baik dari pihak ibu maupun janinnya yang mempunyai dampak tidak baik bagi ibu maupun janinnya. Pihak medis atau tenaga medis akan lebih berhati-hati dalam mengambil tindakan penanganannya.

2.2 Sistem Pendukung Keputusan

Dalam perkembangan zaman saat ini semua aspek ilmu yang bertujuan untuk memudahkan dan membantu pekerjaan manusia dikerahkan dan saling bersaing. Salah satunya sistem yang berbasis komputer bertujuan untuk mendukung sebuah keputusan yang berupa penilaian seseorang dalam mengambil sebuah keputusan namun penilaian tersebut tidak mempengaruhi suatu penilaian. Pada penelitian ini melakukan pengolahan suatu sistem yang menggunakan beberapa model yaitu menganalisis data, penilaian, pandangan untuk menentukan keputusan, dengan pemrosesan berbasis teknologi dengan tujuan sebagai pendukung dalam mengambil sebuah keputusan dengan tertata [7].

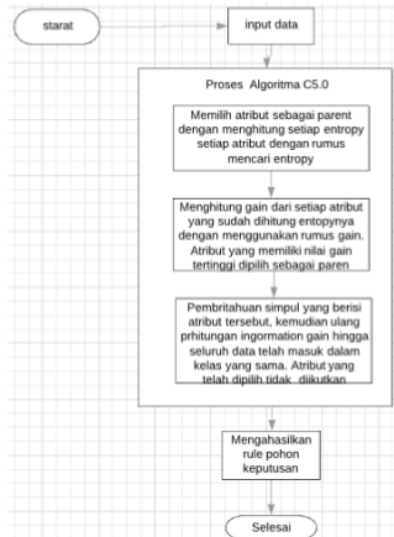
18

2.3 Algoritma C5.0

Algoritma C5.0 merupakan algoritma pembaharuan dari algoritma C4.5 dan sering digunakan untuk dalam pengolahan data berbasis data mining. Kelebihan algoritma C5.0 pendahulunya dalam segi memori memiliki kecepatan memori mencapai 90% daripada algoritma pendahulunya yaitu C4.5 [6], dan dari segi pemakaian memori algoritma C5.0 ini menggunakan memori lebih kecil, contohnya dalam pembuatan rule set pada dataset forest, dimana algoritma pendahulunya menggunakan kurang lebih 3GB memori sedangkan algoritma C5.0 memiliki kurang dari 200MB memori. Yang lebih penting lagi dari segi keakuratan algoritma C5.0 mempunyai tingkat kesalahan yang amat rendah dari pendahulunya, dan menghasilkan pohon keputusan yang lebih kecil dan rule set yang sedikit sehingga lebih membantu dalam melakukan

seleksi fitur [13].

Untuk menghasilkan fitur yang telah terseleksi, penerapan algoritma C5.0 punya beberapa tahapan. Seperti yang dapat dilihat pada Gambar 1 berikut ini.



Gambar 1. Pemodelan Algoritma C5.0

Hal pertama yang dilakukan adalah mengidentifikasi sampel dari data set. Kemudian menghitung nilai entropy dari keseluruhan atribut, lalu selanjutnya yaitu menghitung nilai information gain tertinggi dari seluruh atribut sehingga didapatkan atribut yang akan digunakan sebagai akar atau parent. Selanjutnya percabangan pada akar untuk setiap nilainya ditentukan, kemudian setiap cabang berisi kasus yang telah dibagi. Kemudian perhitungan secara berulang dilakukan untuk menentukan nilai gain. Perhitungan tersebut berhenti ketika semua data yang telah dihitung memiliki persamaan pada kelasnya. Berikut tahapan perhitungan entropy dan gain dalam pembentukan pohon keputusan algoritma C5.0.

1. Jumlah kasus Kemudian dari semua atribut yang ada dihitunglah nilai entropy-nya dan setelah itu barulah menghitung information gain.
2. Persamaan 1 berikut merupakan persamaan untuk mencari nilai entropy sebelum dilakukannya perhitungan dalam mencari information gain.

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

3. Persamaan 2 merupakan persamaan untuk menentukan information gain. Kemudian setelah information gain di dapat tentukanlah information gain yang memiliki nilai tertinggi. Itulah yang akan menjadi akar atau parent pada silsilah pohon keputusannya.

$$E(A) = \sum_{j=1}^n \frac{s_{1j} + \dots + s_{mj}}{s} I(s_{1j}, \dots, s_{mj}) \quad (2)$$

Keterangan :

S : Himpunan Kasus

A : Fitur

n : Jumlah Partisi Atribut A

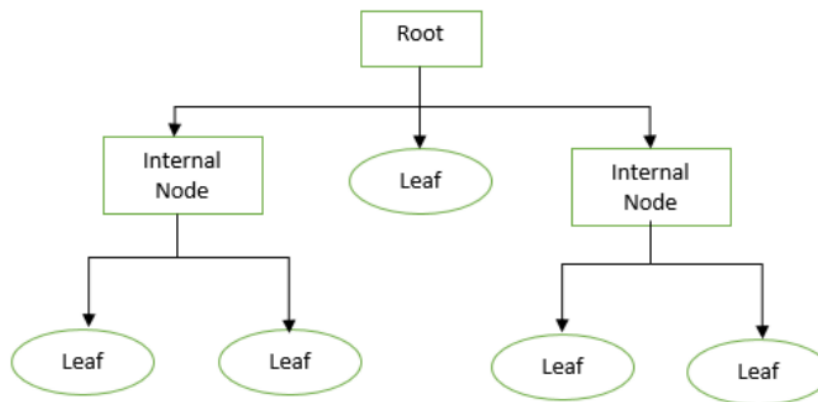
ISI_j : Proporsi S_i Terhadap S

ISI : Jumlah Kasus Dalam S

4. Langkah menentukan information gain di ulang kembali, dan berhenti ketika semua atribut dan subset sudah dihitung.
5. Ketika tidak ada lagi atribut yang dibagi atau di partisi dan keseluruhan relasi dalam simpul N memiliki kelas yang sama maka proses pembentukan pohon dihentikan.

2.4 Decision Tree

Decision Tree atau pohon keputusan ialah salah satu dari sekian banyak metode klasifikasi dengan struktur *tree* yang sama dengan flowchart, dimana masing-masing *node* menunjukkan sebuah hasil pada atribut, yang masing – masing cabangnya mewakili hasil pengujian dengan *node* dan *leaf* memegang label *class*. Dan pada *node* paling atas disebut *root node* seperti Gambar 2 [8].



Gambar 2. Decision Tree (Pohon Keputusan)

Penjelasan lanjut terkait *node* pada metode *decision tree*:

1. **Root Node** ialah *node parent* atau *node* yang paling atas dari pohon keputusan, pada *node* ini tidak bisa dilakukan input dan dapat menghasilkan output dan tidak menghasilkan apa-apa.
2. **Internal Node**, ialah *node* percabangan dari *node* sebelumnya. Pada *node* ini hanya mempunyai satu inputan dan bisa menghasilkan beberapa output.
3. **Leaf Node**, merupakan *node* terakhir. Pada *node* ini hanya mempunyai satu inputan dan tidak mempunyai output.

2.5 Algoritma K-Medoids

K-Medoids atau bisa disebut *Partitioning Around Medoids* (PAM) adalah salah satu dari algoritma *clustering* yang mirip dengan algoritma *K-Means*. Perbedaan dari kedua algoritma ini yaitu algoritma *K-Medoids* atau PAM menggunakan objek sebagai perwakilan (*medoid*) sebagai pusat cluster untuk setiap cluster, sedangkan *K-Means* menggunakan nilai rata-rata (*mean*) sebagai pusat cluster. Kelebihan dari Algoritma *K-Medoids* lebih mudah untuk mengatasi kelemahan yang sering terjadi pada algoritma *K-Means* seperti sensitive terhadap noise dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data. Dan hasil proses clustering dari algoritma *K-Medoids* tidak bergantung pada urutan masuk dari dataset.

2.6 Bahasa Pemrograman R

R merupakan salah satu dari sekian banyak bahasa pemrograman yang digunakan untuk pengolahan data dengan berbasis open source. Dalam pengolahan data secara statistik dan juga untuk menampilkan hasil pengolahan R sering kali menjadi pilihan untuk menghasilkan data dalam bentuk grafik, karena bahasa R sangat mendukung pengolahan data tersebut dikarenakan R memiliki library atau media pendukung yang bisa dipakai dan disajikan ke dalam bentuk *package* sehingga mudah untuk di aplikasikan [11].

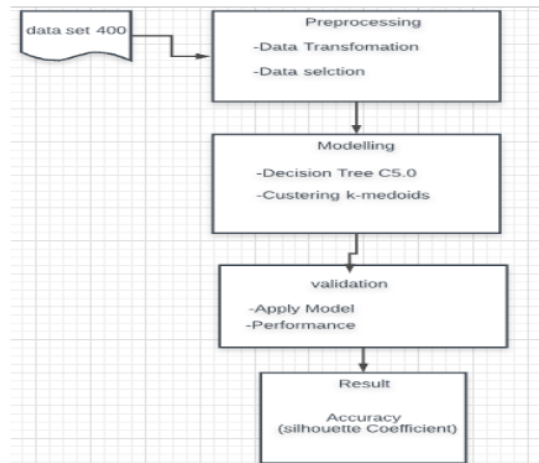
Ciri khas bahasa R terdapat pada bagian *syntax* yang selalu diawali dengan *syntax* ">" dan adapun beberapa keunggulan bahasa R :

1. R sangat unggul dalam segi mengolah data, juga media penyimpanannya yang relatif kecil [11][16].
2. Layanan R yang sangat lengkap dalam mengoperasikan perhitungan array.
3. Dalam penanganan penelitian dalam bidang statistik seperti menguji statistik, menguji fungsi dalam probabilitas dan sebagainya R adalah bahasa yang sangat menunjang dalam bidang statistik.
4. Dari segi tampilan grafik perangkat lunak R menyediakan tampilan grafik yang menarik bagi penggunaannya dan juga fleksibel untuk digunakan.

Adapun fungsi R yaitu multiplatform, dimana multiplatform tersebut memiliki arti R dapat dengan mudah menyesuaikan di berbagai sistem operasi, tidak hanya satu jenis sistem operasi saja. R dapat diinstall di berbagai sistem operasi seperti windows, LINUX dan juga dari versi Macintosh [16].

3. Metodologi Penelitian

Sebelum melakukan modeling data yang digunakan pada penelitian ini sudah melalui proses seleksi data dan transformasi data, untuk melakukan pengukuran dalam penelitian ini menggunakan tool R. Diagram alir tahapan untuk menemukan akurasi yang didapat, pada Gambar 3.



Gambar 3. Proses Metode

Dalam pengujian data set pertama dilakukan preprocessing data agar memudahkan sistem dalam pengolahan data. Adapun tahapan selanjutnya modelling dengan menggunakan decision tree untuk melakukan seleksi atribut dimana dari keseluruhan atribut akan diambil atribut yang paling berpotensi.

Atribut tersebut akan diklusterkan menggunakan metode *k-Medoids* dengan pembagian 10 cluster secara acak dan akan dilakukan pengujian menggunakan *Silhouette Coefficient*. Seluruh tahapan tersebut akan diimplementasikan dalam bahasa pemrograman R dengan pemodelan seperti berikut.

3.1 Pemodelan Algoritma K-Medoids

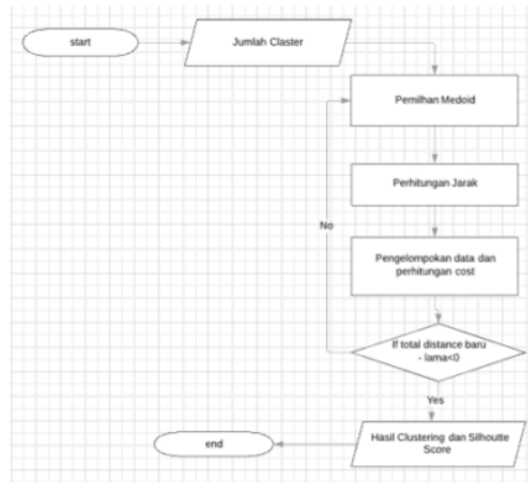
Pada tahap ini dilakukan implementasi dengan langkah-langkah sebagai berikut [2]:

1. Inisialisasi pusat cluster sebanyak *k* (jumlah cluster). Pada tahap ini dilakukan pemilihan objek, objek yang digunakan disini merupakan data dari kuesioner tracer study dan atribut yang digunakan berjumlah 16. Selanjutnya secara acak sebagai pusat cluster sebanyak jumlah yang ditentukan. Seumpama jumlah cluster yang ditentukan ada 5 maka objek yang digunakan harus sama jumlahnya yaitu 5.

2. Alokasikan 1 setiap 1 data (objek) ke *cluster* terdekat menggunakan persamaan 1 ukuran 1 jarak Euclidian Distance. Pada tahap ini setiap objek dilakukan perhitungan jarak terhadap setiap pusat *cluster* yang sudah ditentukan di awal dengan menggunakan perhitungan jarak Euclidian Distance dengan perhitungan Persamaan 3.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; 1, 2, 3, \dots, n \quad (3)$$

- Dari hasil perhitungan dengan Euclidian Distance didapatkan hasil yang mana setiap objek yang dihitung akan di kelompokkan sesuai kedekatan jarak dengan pusat cluster.
3. Pilih secara acak objek pada masing-masing cluster sebagai kandidat medoid baru. Pada tahap ini dilakukan pemilihan secara acak data objek dari setiap cluster sebagai medoid baru atau pusat cluster yang baru. Sehingga pada tahap ini dilakukan pergantian pusat cluster dengan pusat cluster yang baru sesuai jumlah cluster yang sudah ditentukan pada tahap pertama.
 4. Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoid baru. Pada tahap ini setiap objek dilakukan perhitungan jarak dengan medoid yang baru atau pusat cluster yang baru menggunakan Euclidian Distance. Sehingga di dapatkan kelompok atau cluster data yang baru sesuai dengan kedekatan jarak objek terhadap pusat cluster yang baru.
 5. Hitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika $S < 0$, maka tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai medoid. Pada tahap ini dilakukan perhitungan total simpangan antara total jarak objek dengan pusat cluster yang lama dengan total jarak objek dengan pusat cluster yang baru. Selanjutnya dilakukan pengurangan antara total jarak baru dikurangi total jarak yang lama. Jika hasil pengurangan kurang dari 0 (nol) maka dilakukan pergantian pusat cluster dengan pusat cluster yang baru yang sudah dilakukan pada tahap 3.
 6. Ulangi mulai dari langkah 3 sampai 5 hingga tidak terjadi perubahan medoid. Pada tahap ini dilakukan pengulangan proses pada tahap ke 3 sampai ke 5. Jika nilai total proses pengurangan yang dilakukan pada tahap 5 lebih dari 0 (nol) maka proses dihentikan, dikarenakan tidak adanya pergantian medoid atau pusat cluster yang baru.
 - 7.



Gambar 4. Pemodelan Algoritma K-Medoids

3.2 Tahapan Perhitungan Kualitas Dan Kekuatan Dari cluster

Silhouette coefficient merupakan metode yang digunakan untuk melihat ⁶ kualitas dan kekuatan dari cluster. Metode silhouette coefficient merupakan gabungan dari dua metode yaitu metode cohesion yang berfungsi untuk mengukur seberapa dekat relasi antara objek dalam sebuah cluster, dan metode separation yang berfungsi untuk mengukur seberapa jauh sebuah cluster terpisah dengan cluster lain. Tahapan perhitungan *silhouette coefficient*

1. Hitung rata-rata jarak objek dengan semua objek lain yang berada di dalam satu cluster dengan Persamaan 4.

$$a(i) = \frac{1}{[A]-1} \sum_{j \in A, j \neq i} d(i, j) \quad (4)$$

2. Hitung rata-rata jarak objek dengan semua objek lain yang berada pada cluster lain, kemudian ambil nilai paling minimum dengan Persamaan 5.

$$d(i, C) = \frac{1}{[A]} \sum_{j \in C} d(i, j) \quad (5)$$

3. Hitung nilai silhouette coefficient dengan Persamaan 6.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

Nilai hasil ⁴ silhouette coefficient terletak pada kisaran nilai -1 hingga 1. Semakin nilai *silhouette coefficient* mendekati nilai 1, maka semakin baik pengelompokan data dalam satu cluster. Sebaliknya jika nilai silhouette coefficient mendekati nilai -1, maka semakin buruk pengelompokan data di dalam satu cluster.

4. Implementasi

4.1 Data ibu hamil

Data yang di gunakan pada penelitian ini adalah data ibu hamil yang sedang hamil ataupun ibu yang hamil yang telah melakukan persalinan, data ini berjumlah 400 data dengan atribut sebanyak 114 data. contoh data seperti pada Tabel 1 berikut.

Tabel 1. Data Ibu Hamil

Hamil_ke	UI	¹ komplikas	Ab	ITP	²⁵ FD	²⁴ normal
2	28	tidak	tidak	tidak	tidak	ya
2	25	tidak	tidak	tidak	tidak	ya
2	26	tidak	tidak	tidak	tidak	ya
2	30	tidak	tidak	tidak	tidak	ya
2	33	tidak	tidak	tidak	tidak	ya
2	29	tidak	tidak	tidak	tidak	tidak
2	34	tidak	tidak	tidak	tidak	ya
2	20	tidak	tidak	tidak	tidak	tidak
1	22	tidak	tidak	tidak	tidak	tidak

4.2 Implementasi C5.0

Model yang di dapatkan berisi atribut gejala-gejala apa saja yang dijadikan parent sampai pada bagian leaf. Ini bertujuan untuk mengetahui kecenderungan seperti apa yang di alami ibu yang hamil tersebut yang dapat mengakibatkan kehamilan beresiko tinggi, dapat di klarifikasikan hanya dengan melihat atribut dari hasil algoritma C5.0 sehingga atribut yang awalnya 114 atribut dapat berkurang jumlahnya menjadi beberapa atribut. Setelah menghasilkan fitur atau atribut baru dilakukan pengklusteran dengan menggunakan k-medoids. Ini bertujuan untuk mempermudah penganalisaan dalam kelas tersebut, agar dapat menjadi acuan tenaga kesehatan untuk melakukan penyuluhan agar lebih tepat sasaran.

Untuk menghasilkan fitur yang terseleksi dan juga model maka dilakukan implementasi pada tool R. Source code pada *tool* R dengan menggunakan *library* yang sudah di sediakan seperti pada Gambar 5 berikut.

```
require(C50)

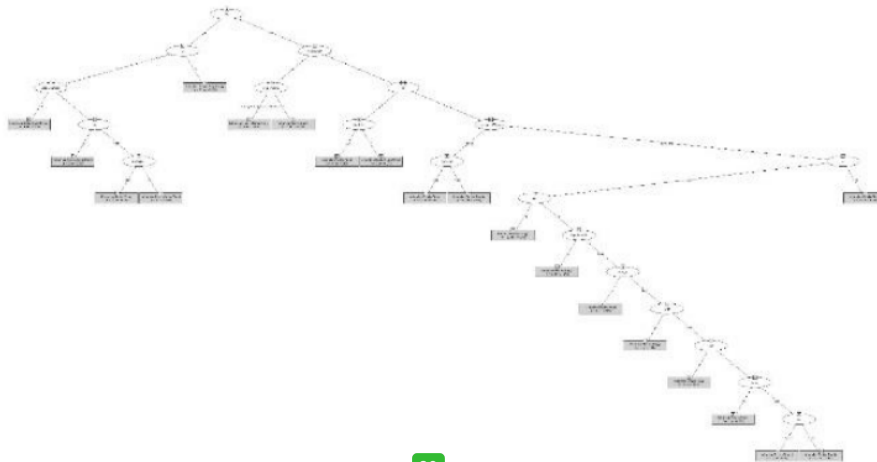
setwd("E:")
data_ibu_hamil <- read.csv("DataIbuHamil.csv")

hasil <- C5.0(data_ibu_hamil[,-21],as.factor(data_ibu_hamil$Kelas))

jpeg("E://hasilgambar.jpg", width = 4096, height = 2048)
plot(hasil, type = "s")
dev.off()
```

Gambar 5. Source Code Pemodalan Algoritma C5.0

Pengujian metode C5.0 menghasilkan pohon keputusan seperti yang terlihat pada Gambar 6 berikut.



Gambar 6. Pohon Keputusan

Dari pohon keputusan tersebut akan dapat terlihat yang mana saja atribut-atribut yang sangat berpengaruh. Maka diambil atribut yang sangat berpengaruh tersebut sebagai atribut prioritas dan di gunakan sebagai fitur baru untuk diklusterkan menggunakan k-medoids berikut atribut beserta data atribut nya dapat di lihat pada Tabel 2.

Tabel 2. Data Fitur Baru

1	Pendarah: UI	LetakJanin	Ab	Hb	Hipertensi	
tidak	tidak	28	tidak	tidak	11.36	tidak
tidak	tidak	25	tidak	tidak	11.8	tidak
1	tidak	26	tidak	tidak	11.8	tidak
tidak	tidak	30	tidak	tidak	10.38	tidak
tidak	tidak	33	tidak	tidak	9.18	tidak
ya	tidak	29	tidak	tidak	11.8	tidak
tidak	tidak	34	tidak	tidak	9.28	tidak
ya	tidak	20	tidak	tidak	10.48	tidak
tidak	tidak	22	tidak	tidak	11.16	tidak

4.3 Implementasi K-Medoids

Data lengkap pada table 4.1 dapat dilihat pada lampiran. Data fitur baru digunakan pada penelitian ini berjumlah 400 data dengan 20 atribut dan diklasterkan. Untuk menentukan nilai cluster yang terbaik dilakukan lah pengujian sebanyak tiga kali, untuk mengetahui nilai *silhouette* ²³ *efficient* mendekati nilai 1. Adapun penerapan *K-Medoids* dengan *Source code* pada tool R dapat dilihat pada Gambar 7 berikut

```

1 library(cluster)
2 library(factoextra)
3
4 hum <- pam(x=DataIbuHamilHasil,k=2)
5 hum$clustering
6 hum$medoids
7
8 summary(hum)
9 fviz_cluster(hum)
10

```

Gambar 7. Source Code Pemodelan K-Medoids

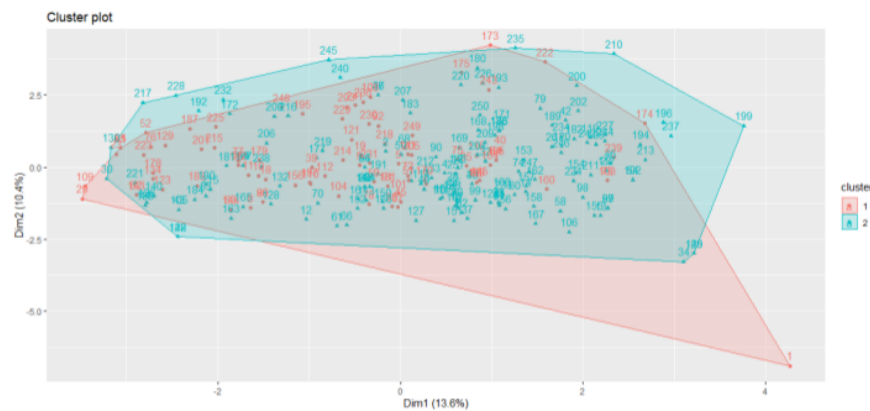
Pada Gambar 7 load package menggunakan `library(cluster)` untuk algoritma *clustering* dan `library(factoextra)` sebagai algoritma clustering dan visualisasi.

4.4 Hasil dan Pembahasan

Untuk menentukan hasil cluster yang terbaik dilakukan pengujian sebanyak 3 kali untuk menentukan nilai k (cluster) yang terbaik.

4.4.1. Pengujian ke-1

²² Pengolahan fitur menggunakan metode K-medoids menghasilkan cluster seperti pada Gambar 8 berikut.



Gambar 8. Hasil Cluster Plot K-Medoids

Klasterisasi pada Gambar 8 menggunakan acuan data dengan ID 230, 58 sebagai pusat cluster. Karakter data dengan ID 230, 58 pada tabel 12 berikut.

Tabel 3. Data Pusat Cluster K-Medoids

ID	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	cluster
230	2	3	13	5	2	85	3	2	2	2	3	18	3	5	2	2	1	3	1	4	1
58	3	3	17	5	2	31	2	2	2	2	3	15	3	5	9	2	4	2	3	1	2

Pada Tabel 3 terdapat ID data acuan dari setiap cluster dan dilakukan perulangan pembuatan sebanyak 19.37983 kali dengan perulangan perpindahan sebanyak 16.44492 kali. Dengan nilai numerical per cluster pada Gambar 9 berikut.

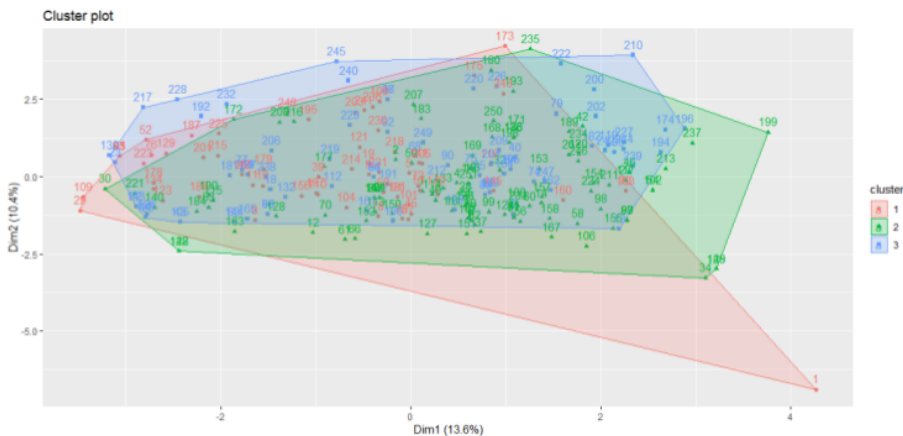
```
Numerical information per cluster:
      size max_diss av_diss diameter separation
[1,]   92 32.35738 17.56060 59.83310   4.690416
[2,]  159 33.98529 15.79936 60.08328   4.690416
```

Gambar 9. Numerical yang Dihasilkan

Pada Gambar 9 nilai numerical yang dihasilkan per cluster dengan cluster pertama memiliki size atribut yang dimiliki sebanyak 92 dengan max_diss mencapai 32.357338 dan av_diss mencapai 17.56060, diameter dari cluster pada cluster pertama mencapai 59.83310 dengan separation 4.690416. Pada cluster kedua memiliki size atribut yang dimiliki sebanyak 159 dengan max_diss mencapai 33.98529 dan av_diss mencapai 15.79936, diameter dari cluster pada cluster pertama mencapai 60.08328 dengan separation sama dengan cluster pertama ini menunjukkan cluster kedua memiliki nilai yang lebih baik dari cluster pertama. Adapun Silhouette Coefficient yang di hasilkan dari kedua cluster pada Gambar 10 berikut.

4.4.2. Pengujian ke-2

Pengolahan fitur menggunakan metode K-medoids menghasilkan kluster seperti pada Gambar 10 berikut.



Gambar 10. Hasil Cluster Plot K-Medoids

Dari cluster p.123 Gambar 10 menggunakan acuan data sebagai pusat cluster. Adapun acuan data tersebut pada Tabel 4.

Tabel 4. Data acuan Cluster K-Medoids

ID	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V16	V17	V18	V19	V20	cluster
142	2	3	11	5	2	96	2	2	2	2	3	11	2	3	2	1	3	3	3	1
250	2	3	15	5	2	26	3	2	3	2	3	13	3	3	2	1	2	3	3	2
209	3	3	17	5	2	48	2	2	2	2	3	16	3	3	2	1	3	1	1	3

Pada Tabel 4 terdapat id data acuan dari setiap cluster dan dilakukan pembuatan pohon keputusan sebanyak 13.64935 kali dengan perulangan sebanyak 13.64935 kali untuk

membentuk ketiga cluster tersebut. Kita juga bias melihat numerical per cluster pada Gambar 11 berikut.

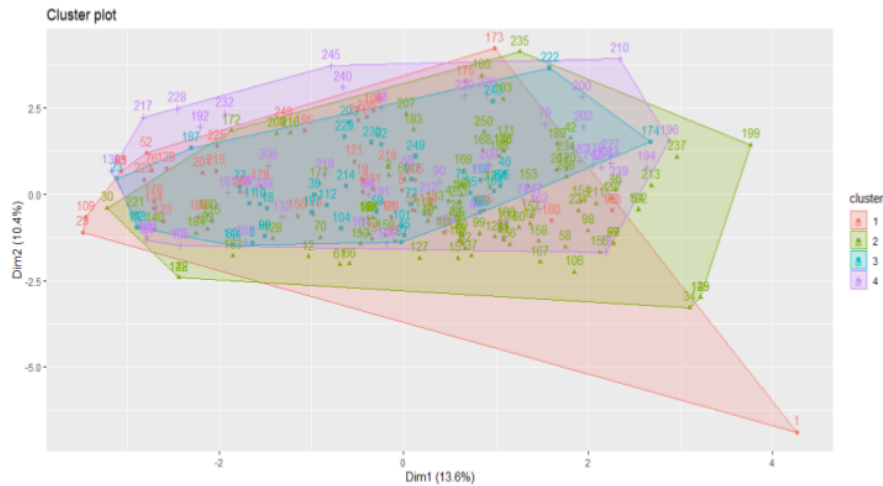
```
Numerical information per cluster:
      size max_diss av_diss diameter separation
[1,]    65 30.93542 13.72167 41.07311  6.324555
[2,]   104 29.63106 13.01927 41.21893  5.196152
[3,]    82 27.12932 14.39113 39.68627  5.196152
```

Gambar 11. Numerical yang Dihasilkan

Pada Gambar 11 nilai numerical yang dihasilkan per cluster dengan cluster pertama memiliki size atribut yang dimiliki sebanyak 65 dengan max_diss mencapai 30.93542 dan av_diss mencapai 13.72167, diameter dari cluster pada cluster pertama mencapai 41.07311 dengan separation 6.324555. Pada cluster kedua memiliki size atribut yang dimiliki sebanyak 104 dengan max_diss mencapai 29.63106 dan av_diss mencapai 13.01927, diameter dari cluster pada cluster pertama mencapai 41.21893 dengan separation 5.196152. Pada cluster ketiga memiliki size atribut yang dimiliki sebanyak 85 dengan max_diss mencapai 27.12932 dan av_diss mencapai 14.39113, diameter dari cluster pada cluster pertama mencapai 39.68627 dengan separation 5.196152.

4.4.3. Pengujian ke-3

Pengolahan fitur menggunakan metode K-medoids menghasilkan kluster seperti pada Gambar 12 berikut.



Gambar 12. Hasil Cluster Plot K-Medoids

Tabel 5. Data acuan Cluster K-Medoids

ID	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	cluster
205	2	3	11	5	2	96	3	2	2	2	3	13	3	3	7	2	3	3	3	4	1
250	2	3	15	5	2	26	3	2	3	2	3	13	3	3	8	2	1	2	3	4	2
45	2	3	15	4	2	73	2	2	2	2	3	16	2	3	5	2	4	2	3	1	
209	3	3	17	5	2	48	2	2	2	2	3	16	3	3	8	2	1	3	1	4	3

Pada Gambar 8 terdapat ID data 205, 250, 45, 209 sebagai acuan dari setiap cluster dan dilakukan pembuatan cluster sebanyak 12.21626 kali dengan perulangan sebanyak 12.12467 kali. Kita juga bias melihat jarak per cluster pada Gambar 12. Pada keempat cluster pada pengujian-3 memiliki nilai *numerical per cluster* pada Gambar 13.

```
Numerical information per cluster:
      size max_diss av_diss diameter separation
[1,]    47 24.00000 11.37182 40.82891   7.141428
[2,]   104 29.63106 13.01927 41.21893   5.196152
[3,]    44 20.92845 11.82182 36.42801   7.141428
[4,]    56 17.88854 11.33309 29.81610   5.196152
```

Gambar 13. Numerical yang Dihasilkan

Pada Gambar 12 nilai numerical yang dihasilkan per cluster dengan cluster pertama memiliki size atribut yang dimiliki sebanyak 47 dengan max_diss mencapai 24.00000 dan av_diss mencapai 13.01927, diameter dari cluster pada cluster pertama mencapai 41.21893 dengan separation 5.196152. Pada cluster kedua memiliki size atribut yang dimiliki sebanyak 104 dengan max_diss mencapai 29.63106 dan av_diss mencapai 13.01927, diameter dari cluster pada cluster pertama mencapai 41.21893 dengan separation 5.196152. Pada cluster ketiga memiliki size atribut yang dimiliki sebanyak 44 dengan max_diss mencapai 20.92845 dan av_diss mencapai 11.82182, diameter dari cluster pada cluster pertama mencapai 36.42801 dengan separation 7.141428. Pada cluster ketiga memiliki size atribut yang dimiliki sebanyak 56 dengan max_diss mencapai 17.88854 dan av_diss mencapai 11.33309, diameter dari cluster pada cluster pertama mencapai 29.81610 dengan separation 5.196152.

Adapun Silhouette Coefficient akurasi per cluster pada pengujian pertama sampai pada pengujian ketiga, seperti pada Gambar 14, Gambar 15, dan Gambar 16.

```
Average silhouette width per cluster:
[1] 0.5551843 0.5926511
Average silhouette width of total data set:
[1] 0.5789182
```

Gambar 14. Silhouette Width Percluster pada Pengujian-1

```
Average silhouette width per cluster:
[1] 0.5616072 0.4415036 0.3051162
Average silhouette width of total data set:
[1] 0.4280493
```

Gambar 15. Silhouette Width Percluster pada Pengujian-1

```
Average silhouette width per cluster:
[1] 0.4898397 0.3112174 0.3081057 0.3581483
Average silhouette width of total data set:
[1] 0.3545898
```

Gambar 16. Silhouette Width Percluster pada Pengujian-3

Dari hasil pengujian 1 sampai dengan 3 dapat dilihat pengujian pertama dengan nilai silhouette paling tinggi dan mendekati 1. Maka pengujian pertama inilah yang nantinya akan membantu dalam menganalisa karakter dari ibu hamil yang berpotensi mengalami resiko kehamilan.

5. Kesimpulan

Dari hasil yang telah ada pada akhir perancangan sistem ini yang menerapkan algoritma C5.0 dan K-Medoids dalam pengolahan data keputusan untuk membantu bidang kesehatan dalam menganalisa dan menangani resiko kehamilan pada ibu, sebagai bentuk pencegahan sebagai berikut :

1. Melakukan perhitungan data menggunakan metode C5.0 dan K-Medoids sebagai acuan penanganan dalam menghadapi resiko kehamilan pada ibu dan calon ibu.
2. Algoritma C5.0 dapat digunakan untuk menentukan fitur yang berpengaruh dalam penanganan kesehatan ibu hamil.
3. Penerapan algoritma K-Medoids sangat membantu dalam penentuan keputusan dalam kecenderungan- kecenderungan yang terjadi pada ibu hamil.

- Dari hasil analisa menggunakan algoritma K-Medoids dengan melakukan pembagian cluster sebanyak 2 cluster dan dilakukan perulangan pembuatan sebanyak 19.37983 kali dengan perulangan perpindahan sebanyak 16.44492 kali mendapatkan nilai rata-rata *Silhouette Coefficient* 0.5789182.

- [1] Kementerian Kesehatan RI. Sistem Kesehatan Nasional: Bentuk dan cara penyelenggaraan pembangunan kesehatan. Pusat Kajian Pembangunan Kesehatan. Jakarta. Departemen Kesehatan. 2009; p.1
- [2] Kementerian Kesehatan RI. Sistem Kesehatan Nasional: Bentuk dan cara penyelenggaraan pembangunan kesehatan. Jakarta: Pusat Kajian Pembangunan Kesehatan Departemen Kesehatan. 2009.
- [3] Prasetyawati AE. Ilmu kesehatan masyarakat untuk kebidanan holistik (Integrasi community oriented ke family oriented). Yogyakarta: Nuha Medika. 2011. p.17-35.
- [4] Kementerian Kesehatan RI. Sistem Kesehatan Nasional: Bentuk dan cara penyelenggaraan pembangunan kesehatan. Jakarta: Pusat Kajian Pembangunan Kesehatan Departemen Kesehatan. 2009.
- [5] Unicef. Ringkasan kajian kesehatan ibu dan anak. 2012. [cited 2018 Agt 28];1. Available at http://www.unicef.org/indonesia/id/A5_B_Ringkasan_Kajian_Kesehatan_REV.pdf
- [6] Tribunnews. Bayi Lahir Saat Usia Kehamilan Kurang Dari 37 Minggu Tergolong Rumit dalam Perawatan. 2017. [cited 2018 5 Sep 03];1. Available at <http://www.tribunnews.com/kesehatan/2017/11/15/bayi-lahir-saat-usia-kehamilan-kurang-dari-37-minggu-tergolong-rumit-dalam-perawatan>.
- [7] Viva. Penyebab Tingginya Angka Kematia Ibu dan Bayi di Indonesia. 2017. [cited 2018 Sep 03];1. Available at <https://www.viva.co.id/gaya-hidup/kesehatan-intim/892600-penyebab-tingginya-angka-kemat-ibu-dan-bayi-di-indonesia>
- [8] Huda LN. Hubungan status reproduksi, status kesehatan, akses pelayanan kesehatan, dengan komplikasi obstetri di Banda Sakti, Lhokseumawe tahun 2005. Jurnal Kesmas. 2007;1(6):275-81.
- [9] Rahmadewi, Herartri R. Faktor-faktor yang berhubungan dengan kehamilan risiko tinggi. Gizi Indon. 2011; 34(2):120-8.
- [10] Nuqson Masykur Huda, 2010, "Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa", Semarang
- [11] Mujib Ridwan, Hadi Suyono & M. Sarosa, 2013, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier", Malang.

Implementasi Algoritma C5.0 Dan K-Medoids

ORIGINALITY REPORT

12%

SIMILARITY INDEX

8%

INTERNET SOURCES

3%

PUBLICATIONS

10%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Unika Soegijapranata

Student Paper

1%

2

ejurnal.stmik-budidarma.ac.id

Internet Source

1%

3

library.umn.ac.id

Internet Source

1%

4

Submitted to Universitas Brawijaya

Student Paper

1%

5

www.tribunnews.com

Internet Source

1%

6

nopi-en.blogspot.com

Internet Source

1%

7

publikasi.dinus.ac.id

Internet Source

1%

8

de.scribd.com

Internet Source

1%

9

Submitted to University of Bristol

Student Paper

<1%

10

Susanti Tria Jaya, Vide Bahtera Dinastiti.
"Pendidikan Kesehatan Kehamilan Resiko
Tinggi Di Desa Ringinpitu Kecamatan
Plemahan", Journal of Community Engagement
in Health, 2020

Publication

<1 %

11

Farid Fadli, Belsana Butar Butar. "Penerapan
Decision Tree Menggunakan Algoritma C4.5
Untuk Deteksi Demam Berdarah Pada RS. IMC
Bintaro", Indonesian Journal on Software
Engineering (IJSE), 2019

Publication

<1 %

12

Submitted to UIN Sultan Syarif Kasim Riau

Student Paper

<1 %

13

eprints.poltekkesjogja.ac.id

Internet Source

<1 %

14

es.scribd.com

Internet Source

<1 %

15

Herlidian Putri. "Determinant Of Preeclampsia
Occurrence In Kalisat Hospital Jember", Jurnal
Kesehatan dr. Soebandi, 2018

Publication

<1 %

16

Submitted to Surabaya University

Student Paper

<1 %

17

mathematics.dk

Internet Source

<1 %

18	adoc.tips Internet Source	<1 %
19	jurnalmahasiswa.unesa.ac.id Internet Source	<1 %
20	lppa-hki.unair.ac.id Internet Source	<1 %
21	www.coursehero.com Internet Source	<1 %
22	Submitted to Universitas Negeri Padang Student Paper	<1 %
23	pt.scribd.com Internet Source	<1 %
24	Submitted to Fakultas Ekonomi Universitas Indonesia Student Paper	<1 %
25	edoc.site Internet Source	<1 %
26	id.scribd.com Internet Source	<1 %
27	Submitted to Sriwijaya University Student Paper	<1 %
28	www.viva.co.id Internet Source	<1 %

29

Febri Hadi. "Penerapan Metode Algoritma C4.5 dalam Menganalisa Pegajuan Kredit pada Koperasi Jasa Keuangan Syariah Kelurahan Limau Manis Selatan", Indonesian Journal of Computer Science, 2018

Publication

<1%

30

Submitted to iGroup

Student Paper

<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off